B1 Cont

"Video Compression and Decompression Processing and Processors" (U.S. patent 5,379,351). These patents are incorporated herein by reference.

At page 8, lines 24 - 31, and page 9, lines 1 – 5, please replace the paragraph as follows (changes are shown on the pages attached hereto):

B2

It will be appreciated that the double-talk detector 318 receives the transmit audio signal on line 342 after the echo has been canceled. This is because it is desirable to compare the received audio signal to the transmit audio signal without the echo. In the case where there is a strong coupling between the speaker 322 and microphone 320 it may be difficult to determine the proper time at which to adjust the filter coefficients. An example scenario is where the speaker is placed near the microphone, and the filter is not yet converged. If there is silence at the near-end, and a far-end audio signal is received (where "far-end" refers to signals received by codec 324), the conditions are proper to adapt the filter. However, the double-talk detector will erroneously detect a near-end signal because the far-end signal fed back to the microphone is not canceled by the echo-cancellation circuitry. When the speaker and microphone are placed near one another, the double-talk detector may never find that it is appropriate to adapt the coefficients, and therefore the coefficients will not converge to a useful state.

At page 13, lines 12 - 30, and page 14, lines 1 – 7, please replace the paragraph as follows (changes are shown on the pages attached hereto):

B3

For video-assisted double-talk detection, the estimated near-end energy, E_near, is combined with the mouth motion energy, E_motion, to calculate the probability of near-end silence P(silence|E_near, E_motion). This is accomplished by calculating, according to the Bayes' Rule:

P(silence|E_near, E_motion) =

P(E_near|silence)*P(E_motion|silence) * P(silence)/(P(E_near)*P(E_motion))

P(E_near|silence) is the probability of observing the particular value of E_near in the case of near-end silence. These values are measured by a histogram technique prior to the operation of the system and stored in a look-up table. P(silence) is the probability of near-end silence and is

usually set to 1/2. P(E_near) is the probability of observing the particular value of E_near under all operating conditions, i.e., both with near-end silence AND near-end speech. These values are also measured by a histogram technique prior to the operation of the system and stored in a look-up table. In the same way, P(E_motion|silence) and P(E_motion) are measured prior to operation of the system and stored in additional look-up tables. In a refined version of the double-talk detector, the tables for P(E_near|silence) and P(E_near) are replaced by multiple tables for different levels of the estimated values of ERLE. In this way, the different reliability levels for estimating E_near in different states of convergence of filter 314 can be taken into account. The resulting probability P(silence|E_near, E_motion) is finally compared to a threshold to decide whether the condition of near-end silence is fulfilled that would allow a reliable, fast adaptation of the filter 314 by adapter 316. In addition, the double-talk detector compares the short-term received audio energy E_receive with another threshold to determine whether there is enough energy for reliable adaptation. If both thresholds are exceeded, an adaptation with a non-zero step-size by adapter 316 is enabled; otherwise the step-size is set to zero.

B3 CONT

At page 15, lines 3 - 17, please replace the paragraph as follows (changes are shown on the pages attached hereto):

In yet another embodiment, the absence of detected mouth movement can be used to advantageously increase the video quality. For example, the hearing impaired may use videoconferencing arrangements for communicating with sign language. Because sign language uses hand movement instead of sound, the channel devoted to audio may instead be used to increase the video frame rate, thereby enhancing the quality of sign language transmitted via videoconferencing. Thus, if no mouth movement is detected, the system may automatically make the necessary adjustments. A related patent is U.S. Patent No. 6,404,776 issued on June 11, 2002, entitled "Data Processor Having Controlled Scalable Input Data Source and Method Thereof," docket number 8X8S.15USI1, which is hereby incorporated by reference. Other embodiments are contemplated as set forth in co-pending U.S. Patent No. 6,124,882 issued on September 26, 2000, entitled "Videocommunicating Apparatus and Method Therefor" by Voois et al., as well as various video communicating circuit arrangements and products, and their

B4